Abstract

Selecting which sub-sequences in a database of nucleic acid such as 16S rRNA are highly characteristic of particular groupings of bacteria, microorganisms, fungi, etc. on a substantially phylogenetic tree. Also applicable to viruses comprising viral genomic RNA or DNA. A catalogue of highly characteristic sequences identified by this method is assembled to establish the genetic identity of an unknown organism. The characteristic sequences are used to design nucleic acid hybridization probes that include the characteristic sequence or its complement , or are derived from one or more characteristic sequences. A plurality of these characteristic sequences is used in hybridization to determine the phylogenetic tree position of the organism(s) in a sample. Those target organisms represented in the original sequence database and sufficient characteristic sequences can identify to the species or subspecies level. Oligonucleotide arrays of many probes are especially preferred.. A hybridization signal can comprise fluorescence, chemiluminescence,. or isotopic labeling, etc.; or sequences in a sample can be detected by direct means, e.g. mass spectrometry. The method's characteristic sequences can also be used to design specific PCR primers. The method uniquely identifies the phylogenetic affinity of an unknown organism without requiring prior knowledge of what is present in the sample. Even if the organism has not been previously encountered, the method still provides useful information about which phylogenetic tree bifurcation nodes encompass the organism.